United States Department of Agriculture

Service Center Initiative (SCI)

Standard for Geospatial Data

Prepared by
Data Management Team #5: Geospatial Data Standards

Abstract: This standard provides the USDA Service Center Initiative with a geospatial data model and data standards. It describes a basic, nationally consistent set of core geospatial data that will provide a foundation on which to base business applications.

Keywords: geospatial data, GIS, metadata, standard

Published and distributed on January 4, 2000
(Revision of Standard for Geospatial Data, SCI Std 007-1, December 17, 1999)

Document number: SCI Std 007-02

Introduction

As directed by the Secretary of Agriculture's March 16, 1998 memorandum, the Natural Resources Conservation Service (NRCS), Farm Service Agency (FSA), and Rural Development (RD) agencies are co-locating offices, modernizing business processes, and partnering to achieve a "one-stop service" for USDA customers at their county-based field offices (Service Centers). One of the major components of the modernization initiative involves the implementation of a Geographic Information System (GIS) across each of the Partner Agencies and in all 2,550 Service Center offices. A Service Center Data Team has been chartered with the overall responsibility for implementing an infrastructure for management of data resources for the Partner Agencies. The GIS Standards Team 5 was formed to address specific data management issues regarding geospatial data.

This version of the standard incorporates revisions to the standard Geospatial Datasets Categories.

The individuals who contributed to the development of this standard are:

David Anderson, (NRCS) Service Center Data Team Leader
Carol Ernst, (FSA) Co-Leader
Emil Horvath, (NRCS) Co-Leader

Liz Cook (NRCS)	Randy Frosh (Unisys)	Nicole Soltyka(SAIC)
Randy English (NRCS)	Tom McCarty (SAIC)	Lee Stieger (FSA)
David Feld (RD)	Jill Schuler (NRCS)	Kent Williams (FSA)

Figure 1—Working group list

Table of contents

This page is intentionally left blank

STANDARD FOR GEOSPATIAL DATA

1. Overview

The objectives of this Standard are to establish and manage the United States Department of Agriculture (USDA) Service Center Initiative (SCI) geospatial data standards; to support the concurrent USDA Service Center Strategy to develop a basic, nationally consistent set of core geospatial data that will provide a foundation on which to base business applications; and to relate to other SCI geospatial standards including the *Standard for Geospatial Dataset Metadata* [A1], the *Standard for Geospatial Feature Metadata* [A2], and the *Standard for Geospatial Dataset File Naming* [A3].

1.1. Scope

The scope of this Standard is to define the minimum set of geospatial data standards required to support the geospatial data layers first defined in the *USDA Service Center Geographic Information System (GIS) Strategy* [A4], and to provide the standards, policies and procedures for acquiring, integrating and delivering minimum and "best available" geospatial data within the USDA enterprise GIS.

1.2. Purpose

The GIS for the Service Center is expected to comprise nationwide coverage of more than 20 common *geospatial datasets* (a group of similar spatial phenomena) collected and maintained at the county level of geography. Four of the *geospatial datasets* are considered critical: ortho imagery, soils, common land units, and census (demographic) data. Some common *geospatial datasets* exist within Partner Agencies or other federal, state, and local sources, while other *geospatial datasets* will need to be generated for the SCI.

Data obtained from agencies and/or other external sources is expected to be available in a multitude of formats and conform to a variety of standards. Due to this situation, this Standard provides a basic set of geospatial standards to which all datasets should adhere to the extent possible. This document describes this basic set of geospatial standards. While these standards will continue to evolve, this document is an initial attempt to identify categories for which standards are needed and to establish initial guidance and direction for Business Process Reengineering (BPR) projects that require geospatial data.

Nationally fielded applications will be developed that rely on the nationally consistent set of geospatial data. These applications will rely on the integrity of the data and that the data meets the specifications in this Standard. Applications that are built locally for a Service Center, or for data that is acquired locally, may require additional or different standards.

1.3. Definitions

Partner Agencies	The three (3) major USDA field service agencies involved in the Service Centers, Natural Resources Conservation Service (NRCS), Farm Service Agency (FSA), and Rural Development (RD).

1.4. Acronyms and abbreviations

APFO	Aerial Photography Field Office
BPR	Business Process Reengineering
CCE	Common Computing Environment
CCITT	Consultative Committee on International Telephone and Telegraphy
CLU	Common Land Unit
DEM	Digital Elevation Models
DOQ	Digital Ortho photo Quadrangle
DRG	Digital Raster Graphics
EPA	Environmental Protection Agency
FEMA	Federal Emergency Management Agency
FGDC	Federal Geographic Data Committee
FSA	Farm Service Agency
FTP	File Transfer Protocol
GeoTIFF	Geo Tagged Image File Format
GIS	Geographic Information System
GNIS	Geographic Names Information System
GPS	Global Positioning Systems
LZW	Limpel-Ziv-Welsh
MDOQ	Mosaicked Digital Ortho Quadrangles
MLRA	Major Land Resource Areas
MrSID	Multi-resolution Seamless Image Database
NMAS	National Map Accuracy Standards
NRCS	Natural Resources Conservation Service
PLSS	Public Land Survey System
RD	Rural Development
SCDT	Service Center Data Team
SCI	Service Center Initiative
SDTS	Spatial Data Transfer Standard
SDE	Spatial Database Engine
SSURGO	Soil Survey Geographic Database
TIFF	Tagged Image File Format
USDA	United States Department of Agriculture
USGS	United States Geological Survey
UTM	Universal Transverse Mercator

2. Geospatial data model

A hierarchical classification will be used to categorize geospatial data. At a detailed level this represents a model that can be used to identify and describe geospatial data in a consistent way. *Geospatial dataset collections*, the highest level of the model, are broken down further into *geospatial dataset categories*, then into *geospatial datasets*, and then into *features*. *Features* are comprised of *geometry, attributes*, *symbology*, and *labels*. The hierarchy is defined as follows:

2.1. Geospatial dataset collection

A Geospatial dataset collection is a catalog and physical repository of geospatial datasets. For example, a USDA data mart that serves geospatial data to only one Service Center, or a USDA national data warehouse that serves geospatial data to all Service Centers.

2.2. Geospatial dataset category

A Geospatial dataset category is a logical group or division of a geospatial dataset collection that includes at least one geospatial dataset complete with all metadata and feature data including geometry, attributes, labels and symbology. For example, the geospatial dataset category *soils* includes Soil Survey Geographic Database (SSURGO) polygons, soil lines, soil points, soil attributes database, MLRA polygons, soil legend files and all metadata.

2.3. Geospatial dataset

A Geospatial dataset is a group of similar spatial phenomena in a geospatial dataset category and is often referred to as a layer, theme, coverage, or simply a map. Each geospatial dataset is related to one metadata set. For example, Figure 1 depicts the geospatial dataset category *hydrography* that contains the geospatial *datasets surface water, water control infrastructure* and *hydrologic unit*. The surface water geospatial dataset contains features such as streams represented as lines, ponds represented as polygons, and wells represented as points.

2.4. Feature

A feature is a point, line, area (polygon), text, raster or grid in a geospatial dataset. A feature includes geometry, topology (if supported), attributes (geospatial and tabular), symbology and labels. For example, in Figure 1, a feature is an instance of a pond, stream, or well.

2.4.1. Geometry

The coordinates that define the spatial extent, shape and location of a feature. The geometry of a feature is defined by a series of x and y (and sometimes z) coordinates. For example, a point is represented by a single coordinate pair and lines and polygons are represented by a sequence of coordinate pairs that define the shape of the feature. Each coordinate pair is a vertex that marks a change in direction of the line or polygon boundary. For example, the coordinates for a point could be (100,100) and a line could be (0,0 100,100).

2.4.2. Attributes

Attributes are name and value pairs that are characteristics or properties of a feature. For example, see the attributes *uid, type, name, temp* and *quality* in Figure 1. The valid values for a particular attribute may be maintained in a related domain table.

2.4.3. Symbology

Symbology is a means to consistently portray geographic objects on a computer display or hardcopy map. An example of symbology that may be used for streams is a dashed blue line.

2.4.4. Label

A label is a distinctive name used to identify a feature, often the primary attribute or unique identification of a feature. Feature attributes may be used as feature labels. An example is the feature attribute *name* used to label the stream in Figure 1 with the value of *Tuscany*.

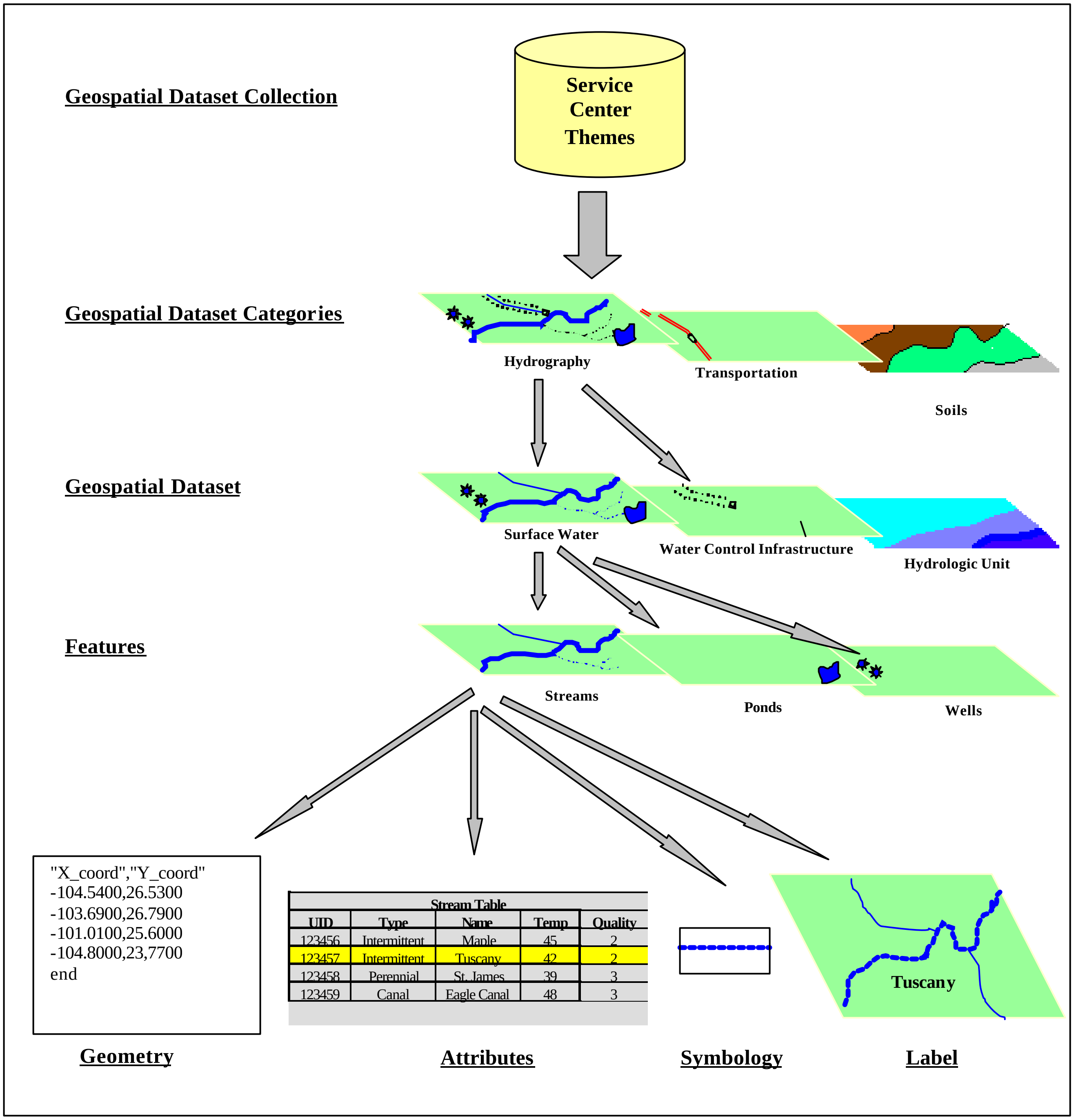

Figure 1 - Geospatial data model

2.5. Geospatial data requirements

A classification or taxonomy of the geospatial data that comprise the SCI geospatial data model is shown below. A business case of a need for this data has been made to administer programs, service customers, and perform distinct enterprise activities. The case for each category and dataset was first made in the USDA Service Center Geographic Information System (GIS) Strategy which identified 19 themes that meet the needs of the business community. The list was modified as data needs were identified and means were realized to use geospatial data to improve the way business is done. The list is detailed in the *Geospatial Data Acquisition, Integration, and Delivery National Implementation Strategy Plan* [A5]. Four themes have consistently been identified as critical. They are ortho imagery, common land unit, soils, and census (demographic).

It is expected that there will be some changes in the list as business needs change, new needs are identified and additional data becomes available. Currently, the list is relatively stable but will likely undergo continual refinement.

There are many geospatial data taxonomies or classification schemes that could be used to characterize the geospatial data requirements for the SCI. Additionally, the list of geospatial data can be divided into three broad categories: framework, natural resources, and business area. The following classification uses terms for geospatial data that are commonly understood within the SCI (bolded terms indicate critical themes):

a) Air Quality
b) Archeology
c) Cadastral
    - Public Land Survey System (PLSS)
    - Military Installations
    - National/State Forests
    - National/State/County Parks
    - Bureau of Land Management Lands
    - Indian Lands
    - Other Ownership
    - Tax Records

d) Census
    - Census Tract Boundaries
    - Census of Population and Housing
    - Census Blocks
    - Census Block Groups
    - Census of Agriculture
    - Economic Census

e) Climate/Precipitation
    - Annual
    - Monthly

f) Climate/Temperature
g) Common Land Unit

h) Conservation Practices
  - Planned and Applied Conservation Practices

i) Elevation
  - Contours (hypsography)
  - Digital Elevation Models

j) Endangered/Habitat

k) Environmental Easements
  - Wetland Reserve Program

l) Geographic Names
  - GNIS Major physical and cultural features (concise)

m) Governmental Units
  - Geographic Names Information System (GNIS) Cities and Towns (populated places)
  - State boundaries
  - County boundaries
  - City boundaries
  - Minor civil divisions
  - Congressional Districts
  - Managed Area Database
  - Soil and Water Conservation Districts
  - Resource Conservation and Development Areas
  - Zip codes

n) Hydrography
  - Environmental Protection Agency (EPA) Reach 3 Files
  - Federal Emergency Management Agency (FEMA) Flood Hazard Maps
  - United States Geological Survey (USGS) Hydrography
  - Water Control Infrastructures/National Inventory of Dams

o) Hydrologic Units
  - 14 digit

p) Imagery

q) Land Use/Land Cover

r) Map Indexes
  - National Aerial Photography Program
  - Quad Index 1:12K, 1:24K, 1:100K, 1:250K

s) Ortho Imagery
  - Digital Ortho Quadrangles
  - County Mosaic

t) Plants

u) Soils
  - SSURGO
  - Soil attribute database
  - Major Land Resource Area (MLRA)

v) Topographic Images
  - Digital Raster Graphs

w) Transportation

- — Roads
- — Railroad
- — Utilities (pipelines, transmission lines, utility lines)

x) Wetlands
- — Agricultural Wetland Determinations (National Food Security Act)
- — National Wetlands Inventory

3. Existing standards

Some of the geospatial datasets are already complete or can be obtained from other sources (e.g., USGS (United States Geological Survey)). In some cases, standards for these data have already been defined and accepted in their respective communities. In these cases, the current standards for these datasets will be assessed for compatibility with Service Center data requirements. If there is a conflict between the current standard, and Service Center requirements, then the Service Center data steward will propose a new or augmented standard to supercede the existing standard.

4. GIS platform

Definition of GIS data standards is, to some extent, dependent upon the GIS platform in use. As such, these standards currently assume that the target Service Center pilot project GIS platform is desktop ArcView®. These standards are subject to change if the GIS platform changes with further definition of the Common Computing Environment (CCE).

5. Base map, map scale, and geospatial accuracy

The Service Center GIS will use both large and small-scale products. However, since digital ortho photo quadrangle (DOQ) coverage is expected to be collected for the entire nation, and is considered a critical foundational geospatial dataset by the Service Center agencies, the preferred map base for the Service Center GIS is the standard DOQ. Collection of new geospatial datasets will be based on the current DOQ where feasible.

Digital map data products in use at the Service Center are expected to be at a scale of 1:24000 or larger. However, smaller map scale (1:50000 or even 1:100000) data may be provided and used at the Service Center if appropriate for the application or when that is the only scale at which data is available. In these cases, users and customers will need to be cautioned that the data may not meet National Map Accuracy Standards (NMAS) which requires that for a 1:24000 scale product, 90% of the locations tested must be within 40 feet (12 meters) of their true location. Where possible, disclaimers will be distributed with the data to discourage the use of data at inappropriate scales. DOQs should not be routinely used for map display or hard copy customer products at scales larger than 1:7920 [1”=660’]. Additionally, users should be aware that many geospatial datasets were digitized from original sources at a scale smaller than 1:24000. This results in mismatches between the digital ortho photo base and the digitized dataset. Over time, producers of these datasets may recompile them on a larger scale source when it is cost beneficial.

Where appropriate, if more accurate Global Positioning Systems (GPS) data does not exist, DOQ data at 1-meter resolution should be used as the base for developing geospatial datasets that do not already exist but fall within the domain of USDA. DOQs, Mosaicked Digital Ortho Quadrangles (MDOQs), and compressed county mosaics produced from DOQs, meet NMAS for the 1:12000 scale, where 90% of the locations tested must be within 33.3 feet (10 meters) of their true location.

To accurately transfer base map information from hard copy to digital, the user should set their "display scale" (if digitizing in ArcView®) to at least match that of the base map scale. The user should be aware that digitizing from a larger display scale than that of the base map may lead to a higher relative accuracy between the source map and the digitized product, but the absolute horizontal accuracy of the digitized product will not be improved.

6. Tiling (coverage unit)

Tiling of digital geospatial data significantly impacts overall data management and system performance. In general, it is preferred that tiling is seamless, or transparent, to the user. However, data tiling schemes may differ between the data warehouse, data mart and local data stores. Studies are currently underway to identify the optimal tiling of digital geospatial data in these various enterprise environments. Table 1 lists examples of various tiling schemes. It assumes the existence of a centrally located national data warehouse that services geospatial data requirements for all state and county offices and of local level data marts that service the geospatial data requirements within the domain of a particular Service Center. This table is for illustrative purposes only and may not represent the actual geospatial data architecture of the SCI.

Table 1–Service center tiling schemes

Data Layer	Location	Tiling Scheme
Digital Ortho photo	National Warehouse	7.5 Minute Quadrangle
Digital Ortho photo	Local Data Mart	Compressed County Mosaic
Soils	National Warehouse	Soil Attribute Database + SSURGO Polygons
Soils	Local Data Mart	By Soil Survey
Hydrologic Units	National Warehouse	Seamless National Layer
Hydrologic Units	Local Data Mart	All Hydrologic Units that Intersect the County

7. Horizontal data integration (edge-matching)

All critical vector geospatial datasets (ortho imagery, soils, common land unit, and census) will be edge-matched across tile boundaries. Both spatial and attribute data must be consistent between tiles. Other common vector datasets could be edge-matched at a

Service Center to the extent determined by cost/benefit. However, if the producing agency has not edge-matched their dataset, it may be delivered to the Service Center without edge-matching. Non-critical raster datasets, Digital Raster Graphics (DRGs) and Digital Elevation Models (DEMs) may be delivered without edge-matching unless the mismatch at map sheet edges do not meet NMAS.

Mismatches between DOQs can be caused by geometric misalignment and/or differing spectral characteristics of the source photography. Mismatches between DOQs will be identified before delivery to a Service Center. Depending on the magnitude of the mismatch, either positionally or visually, DOQs will either be corrected through mosaicking or returned to the source for correction or replacement. According to the *Geospatial Data Acquisition, Integration, and Delivery National Implementation Strategy Plan*, the Aerial Photography Field Office (APFO) will identify mismatches between DOQs and deliver DOQ derived products to the Service Centers. The current plan calls for delivery of a compressed, seamless, tone-balanced county mosaic and uncompressed 7.5-minute tiled MDOQs. Vector datasets properly created using the 7.5-minute DOQ base should not require further edge-matching.

8. Vertical data integration

Vertical integration is the spatial alignment of geospatial objects between layers. For example, vector common land unit (CLU) boundaries matching fence lines visible on an ortho photo image or vegetation boundaries from a land cover layer matching stream lines on a hydrography layer. Since many of the datasets were digitized from sources older than, and at a smaller scale than, the current digital ortho photo base, it is likely that vertical integration will appear poor when "zoomed in" to the data. However, the effort involved in correcting this for some geospatial datasets may be cost prohibitive. Users should be made aware of such situations. Geospatial datasets that are created from the digital ortho photo base should be at least vertically integrated with the ortho photo and with each other.

It has not been determined how current and future versions of the digital ortho photo base will be registered and how this will affect geospatial dataset integration. There are two possible scenarios. In the first scenario, if the newer DOQ base is significantly more accurate than the original DOQ derived base, then the Service Center geospatial dataset created with this base will need to be re-integrated with the newer base, yielding both updated DOQ and feature products. In the second scenario, if the newer DOQ base was not significantly more accurate than the original DOQ derived base, the new DOQ would be registered to the older DOQ base. This situation does not require the re-integration of vector data layers created with the older DOQ base, and yields only the updated DOQ base. The first situation propagates the vertical integration problem while the second diminishes the value of the newer, more accurate DOQ base. All new data collection efforts should be collected with the DOQ as the base map.

9. Topology

All vector data must be topologically clean. This does not necessarily mean that the data will maintain topology, for example ARC/INFO® coverages maintain topology while ArcView® shapefiles can only be topologically clean. Datasets maintained in a database management system and accessed through SDE® (Spatial Database Engine) are based on a continuous non-topological object data model.

Topologically clean implies that vector lines and areas must be free of spatial errors such as slivers and gaps, loop-backs and unnecessary duplicate geometry as well as nodes properly placed at linear intersections. Vertical integration of topology is required for some, but not all vector geospatial datasets. For example, land use polygons and stream lines may require node placement for certain buffering analysis, however, roads and field boundaries may not.

10. Coordinate reference system

All geospatial data (both raster and vector formats) should be represented in a projected coordinate system for use in the Service Centers. The coordinate reference system to be used is the Universal Transverse Mercator (UTM) with the North American Horizontal Datum 1983 and Vertical Datum 1988. Map units of measurement are to be in meters. When a UTM zone cuts through a county, all geospatial data will be projected into the predominant UTM zone for that county.

The coordinate reference system may differ depending on the storage and usage of the dataset. For example, vector data may be more efficiently stored in latitude/longitude coordinates in the data warehouse, while vector data at the local data mart may be more useful projected into the appropriate UTM zone.

11. Attributes

All vector geospatial data will be attributed with at least a primary attribute, or key, such as a unique name or identification number. Additionally, geospatial data may be attributed with a display color and symbology. Geospatial data may have multiple attributes associated with them and if so, these attributes must be formatted and maintained in tables compatible with a relational database system.

A geospatial data model must be provided for each geospatial dataset used in Service Centers. At a minimum, this model shall list for each geospatial dataset: the type(s) of feature(s) (i.e., point, line, area, text, raster or grid), feature geometry, feature attributes, including the primary key field used to link the feature to other tabular databases, and the default feature symbology and feature labels. All attribute data must be defined to the physical level. The SCI Standard for Geospatial Feature Metadata, currently in development, will outline the requirements for documentation of features.

12. Archive format

All warehoused vector data will be archived in a format compatible with the Service Center Initiative enterprise GIS solution. All warehoused raster data will be in non-proprietary, e.g., Geo Tagged Image File Format (GeoTIFF) or Tagged Image File Format (TIFF), either uncompressed or compressed, e.g., Limpel-Ziv-Welsh (LZW), Consultative Committee on Internaltional Telephone and Telegraphy (CCITT) Group 4, or Multi-resolution Seamless Image Database (MrSID). Data at the Service Center or regional data marts will be stored in the most easily accessible native GIS format.

13. Naming

All map/feature type files must be named in accordance with Service Center Data Team (SCDT) standards and must follow any applicable rules and conventions for the GIS platform. The Data Management Team is working on a Standard for Geospatial Dataset File Naming, which outlines the file and directory structure naming conventions for SCI geospatial data and provides guidelines for geospatial data that is obtained from the state or local level.

14. Metadata

A minimum set of digital metadata for each geospatial dataset must be provided. All metadata must conform, at least, to the SCI Standard for Geospatial Dataset Metadata. This Standard is a subset of the FGDC Version 2 Content Standard for Digital Geospatial Metadata. More specifically, data that is produced within USDA for public consumption will have a fully compliant FGDC metadata record when fiscally possible. Data that is produced within USDA that is for Service Center consumption only requires minimum metadata capture. Finally, data that is produced outside USDA, but is for Service Center consumption must have the minimum metadata added if not already present.

Feature metadata must also be generated for each *geospatial dataset*. This includes information about feature geometry, attributes, symbology, and labels. The *Standard for Geospatial Feature Metadata*, currently in development, details the documentation necessary for feature metadata.

15. Distribution

Data will be available in a variety of formats: 1) vector data in a format compatible with the Service Center Initiative enterprise GIS solution or Spatial Data Transfer Standard (SDTS), and 2) raster data in GeoTIFF/TIFF, uncompressed or compressed, or MrSID compressed format) and media (i.e., File Transfer Protocol (FTP), CD-ROM, and 4 mm tar tape).

Appendix A – Bibliography

[A1] Standard for Geospatial Dataset Metadata

[A2] Standard for Geospatial Feature Metadata

[A3] Standard for Geospatial Dataset File Naming

[A4] USDA Service Center Geographic Information System (GIS) Strategy, August 1998

[A5] Geospatial Data Acquisition, Integration, and Delivery National Implementation Strategy Plan, September 1999